Pergamon

Accid. Anal. and Prev., Vol. 29, No. 4, pp. 495-506, 1997

0001-4575/97 $17.00 + 0.00

PII: S0001-4575(97)00029-8

FATIGUE, WORKLOAD AND ADAPTIVE DRIVER SYSTEMS[1]

P. A. HANCOCK[1,*] and W. B. VERWEY[2]

[1]Human Factors Research Laboratory, 141 Mariucci Arena, 1901 Fourth St. S.E., University of Minnesota, Minneapolis MN 55455, U.S.A. and [2]TNO-Human Factors Research Institute, PO Box 23, 3769 ZG Soesterberg, The Netherlands

Abstract—This paper is directed to the further understanding of the problems of fatigue and workload and their role in diminishing driving capability. We present a specific strategy designed to defend against the adverse effects of fatigue and workload extremes through the use of adaptive driver systems. To begin, the work presents a brief critique of Muscio's constraints on developing a test of fatigue. In criticizing these constraints, we point to the commonalities between all energetic reflections of human performance and use advances in stress theory to explain the problems of and resolution to the question of fatigue. We link fatigue and chronic workload and use this coadunation as a basis for the operation of adaptive driver systems which are specifically designed to combat driving impairment. A specific program is then explained in detail and illustrations are given as to how an extension of previous efforts can address the problem of the drowsy and chronically fatigued driver. Future recommendations are articulated. © 1997 Elsevier Science Ltd.

Keywords—Fatigue, Workload, Adaptive systems

THE QUESTION OF FATIGUE

The question has been raised as to how we, as a global society, deal with the problem of fatigue while driving. The answer is quite simple. Don't allow individuals to drive while fatigued! Unfortunately, a capitalist, market-driven economy which reifies efficiency, does not want to hear this answer. So-called, 'real-world' constraints mean that some individuals, especially commercial vehicle drivers, are going to drive fatigued, whether it is advisable to or not. Until we change societal mores, the initial question should actually read, how can we permit individuals to drive while fatigued and not have them punished for the mistakes that are sure to follow. It is true we do legislate some aspects of extended driving[2], however, we cannot legislate others, for example the leisure time activities of drivers, so the process at present is directed to the diminution of untoward effects, rather than the elimination of fatigue, in total.

The problem of what to legislate in order to mitigate fatigue gets straight to the heart of the problem, that of definition. Fatigue researchers have been impeded by an unprofitable contemplation of the Muscio (1921) paradox and it is this paradox then we intend to first resolve. In attempting to develop a measure, Muscio foundered on the problem of definition and his failure has pervaded the research literature in this area since. Muscio posed two conditions for a valid test of fatigue. First, that we know what we mean by fatigue and second, that we have some method, other than the fatigue measure under scrutiny, to know whether fatigue is actually present and in what degree. Essentially, an independent and valid measure against which to compare the proposed fatigue test itself. He inferred from his examination that, based on these constraints, a fatigue test was not possible, a conclusion that has had much influence upon subsequent investigators (see also Broadbent, 1979). While Muscio's conclusions follow from his premises, we claim these premises are fallacious.

If we have to know what any concept means before we conduct substantive research on it, empiri-

*Corresponding author.

[1]This work is dedicated to the memory of Dr. John Hachiya, a good physician, a good man, and a good friend. He is missed by those who knew him.

[2]One obvious example is the number of permissable driving hours and associated rest hours that we impose on commercial drivers. However, regulation does not ensure compliance. The Journal of Commerce (11-8-96) featured an article concerning the prosecution of an owner of a trucking firm for alleged violations of driving hours in the whole company. If convicted he faces a custodial sentence. The article further refers to drivers' log books "known in some parts of the industry as 'joke books'". Consequently, legislation that is hard to enforce and often runs counter to the stimulus-driven human behaviors such as "get homeitis" have a lesser influence than originally hoped by well-meaning legislative authorities.

cist science is lost. Kant's (1787) 'Critique of Pure Reason' is centered upon the question of whether knowledge follows experience or whether knowledge can be distilled from pure contemplation. Contemporary science contends that definition follows upon observation[3]. It is the case that many fundamental physical concepts such as space and time, do not have unequivocal definitions even now. Further, there are many affective states, e.g. attention, of which we still have only a minimal level of understanding, but that has not inhibited research progress on such a topic as it has done in the case of fatigue.

We may be less literal in our interpretation of Muscio's first condition and treat the word 'know' in the sense of an acceptable definition. The implication from his subsequent discussion is that Muscio meant exactly this. Indeed, he notes that "it is not necessary that our knowledge of the nature of fatigue be exhaustive; incomplete knowledge, of a particular kind about fatigue would be sufficient to render the search for a fatigue test reasonable". With this perspective, we can be kinder and more sympathetic to Muscio's position. His dilemma reminds us of Francis Bacon's admonition concerning phenomena that he calls the 'Idols of the Market-Place'. In Bacon's own words:

> The idols (of the market-place) that are imposed by words on the understanding are of two kinds. Either they are names of things that do not exist (for just as there are things without names because they have never been seen, so also there are names without [corresponding] things, as a result of fanciful suppositions); or they are names of objects which do exist but are muddled and vague, and hastily and unjustly derived from things. Words of the first kind are 'fortune', 'prime mover', 'planetary orbs', 'the element of fire' and other fictions of this kind, the product of groundless and false theories. And this class of idols is easier to get rid of, because they can be destroyed by steadfast denial and rejection of the theories. But the other kind is obscure and deep-seated, and is derived from an incorrect and unskilled abstraction. For example, let us take any word, moist, say, and see how far the things which are signified by this word agree with each other; we shall find this word moist is nothing but a confused mark of different actions, which do not allow reduction to any consistent meaning. For it signifies not only something that readily surrounds another body, but also something with no definite boundaries and unable to become solid; something which yields easily in every direction; something which easily subdivides and scatters itself; or easily coalesces and becomes one; easily flows and is set in motion; easily adheres to another body and makes it wet; and which easily liquefies, or melts, when it was previously solid. So when we predicate the word in one sense, a flame is moist, in another, glass is moist. Thus it is easy to see that this notion [of moistness] is carelessly abstracted only from water and common and ordinary liquids, without any proper verification. In words there are, however, certain degrees of wrongness and error. A less faulty class is that of the names of substances, especially of well-derived species of the lowest [type] (for the notion of chalk or mud is good, that of earth is bad). A more defective kind is that of actions like to generate, to corrupt, to alter; the most defective of all is that of qualities (other than those that are the immediate objects of the sense), such as heavy, light, rare, dense, etc. And yet among all of these there are inevitably some notions that are slightly better than others, depending on how many things strike the human sense.

Although it might appear that fatigue is of the latter property, it is the case that most human beings have had sufficient experience with fatigue that we can agree that it is a useful description of a commonly experienced state. In this respect, Muscio's first premise is doubtful in either interpretation. We do not have a closed-end understanding of the concept of fatigue, nor need we have to conduct research investigations. Indeed, under such strictures, scientific inquiry would never begin.

Muscio's second point is similarly flawed. In it, he requires that we already have the very thing that we search for. That is, to provide a test for fatigue, one must first have a test for fatigue. This tautology is most disturbing and it is disconcerting that anyone would consider it grounds for denying the possibility of developing a valid test. On a priori grounds, such a constraint would make the measurement of anything impossible. Apparently acceptable measures of physical entities such as space, are only bootstrapped to artificial definitions. The dimension of time is in a much worse condition, in which the concept is identified with measure itself! Although we have pointed to what we believe are flaws in Muscio's argument, there is much to contemplate in his pivotal work. Many of his points are well made, especially the critical problem of assessment of amorphous concepts. However, in adopting an ultra-rigorous position, research and understanding of fatigue has certainly not progressed to the same extent as developments in other areas of energetic facets of human capability, e.g. stress, workload, attention, etc. (Freeman, 1948).

It would be wrong to lay the lack of progress solely at Muscio's door. Indeed, we should note that it is Muscio who first points to many of the critical problems to be solved in the search for fatigue assessment. There are a spectrum of problems specific to the area, which plague fatigue research. However, if fatigue is seen as a close relation of stress, we can take recent progress in conceptualizing stress effects to help understand fatigue.

THE SOLUTION TO FATIGUE

Fatigue, like stress, suffers from the locus of description problem. That is, in our search for caus-

[3]This is not to say that we directly agree with the present zeitgeist. Indeed, there is much to be argued for in the case of pure reason.

ality, we locate the causal agents of fatigue either in the environment, in the individual themselves and in the interaction between the two. For example, we often seek to understand how hours at work and time-of-day influence fatigue (Monk and Folkard, 1983). Others search for subjective and physiological correlates of the fatigue state, while yet another group of investigators examine changes in performance. In stress research, we (Hancock and Warm, 1989) have labeled these different reflections as the 'trinity of stress', which are input, adaptation and output. Like stress, we can view fatigue as induced by some aspects of the input to the human from the environment, e.g. heat, noise, work hours, etc. In the laboratory, these conditions are perfectly repeatable and so the 'input' agents are largely deterministic (although, of course, the real-world never repeats itself). This approach is embraced by researchers whose orientation is for physical causal models. Their frustration is always reflected in the variability and difference in individual response.

For those who adopt an adaptation model of fatigue, the search is for response mechanisms which represent the adjustments made by the individual to these precursory input agents to fatigue, such as long hours at work. These researchers have a particularly difficult quest, since fatigue does not produce a systematic change in a single physiological parameter, nor, at present does it seem to produce systematic changes in a matrix of physiological responses, although that search continues in earnest. This search, however, is most seductive since it promises to link measurable change in known precursory conditions to adjustment mechanisms, which can then be linked to performance change. It promises, therefore, to present a complete description of the fatigue sequence. As we indicate below, there is some reason to hope this is not an empty promise.

Finally, other researchers concern themselves directly with change in performance. At an extremely pragmatic level, it might be observed that we do not really care what causes degradation in performance, only in what can correct potentially problematic response. So, the source of driver deterioration, e.g. fatigue, drunk driving, epilepsy, etc., is not the direct concern, only how a remediative system can be constructed to correct the driver's inappropriate actions. In the case of fatigue, this is usually associated with an oversteer reaction following a startle response from an episode of inattention or incipient sleep. From an output viewpoint, corrective mechanisms such as momentary lock-out on excess steering correction might be proposed.

In reality, it is rare if anyone concerns themselves solely with only one of these three facets of fatigue. Most researchers looking at environmental effects are directly measuring performance and are seeking appropriate physiological measures which represents the adjustments and effort which accompany imposed fatigue. Even those directly responsible for on-the-road safety, search for underlying causal mechanisms in their efforts to provide and legislate a better driving environment. Consequently, it is our conclusion that fatigue is a multi-faceted entity that needs to be understood from each respective position and the way we intend to clarify this is with reference to the theory of Hancock and Warm (1989).

Our definition of fatigue is: an individual's multidimensional physiological–cognitive state associated with stimulus repetition which results in prolonged residence beyond a zone of performance comfort. It is important here to unpack this definition and this can best be done in reference to Fig. 1, a complete description of which is given in Hancock and Warm (1989). Briefly, the base axes are information rate (the temporal complexity of information in a specific environment), and information structure (the spatial complexity of the environmental display in respect of meaning to the individual observer). Excess or insufficiency on these axes remove the individual from their optimal capability illustrated at the peak of the figure and degrades their response accordingly. Where any one individual starts on these respective axes as a 'comfortable' region of performance is a function of their personal history. Consequently, how different forms of input affects the functioning individual depends upon genetic and learned differences.

Information rate and structure represent the input values. The degree to which individuals can adjust their reactions to these values represent the adaptation facet of response. Finally, the output is level of performance described by that region of the figure in which the individual resides having engaged in their own respective adaptation to the current input. There are many environmental factors which can induce the fatigue state. It is possible to express each of these in terms of information rate and structure. For example, typical vigilance experiments in psychology present a very low rate of stimulus presentation and very restricted spatial structure (Warm et al., 1996). We know from work on sensory and perceptual deprivation that prolonged residence in these conditions is highly disturbing (Hancock, 1980). What links vigilance and fatigue here is the stimulus rate which is uniformly low.

Prolonged time at work is also a precursory condition to fatigue. In terms of cognitive adaptation, it is the enforced repetition of the same actions which prove problematic. Eventually, prolonged work also has an influence on physiological reflections of

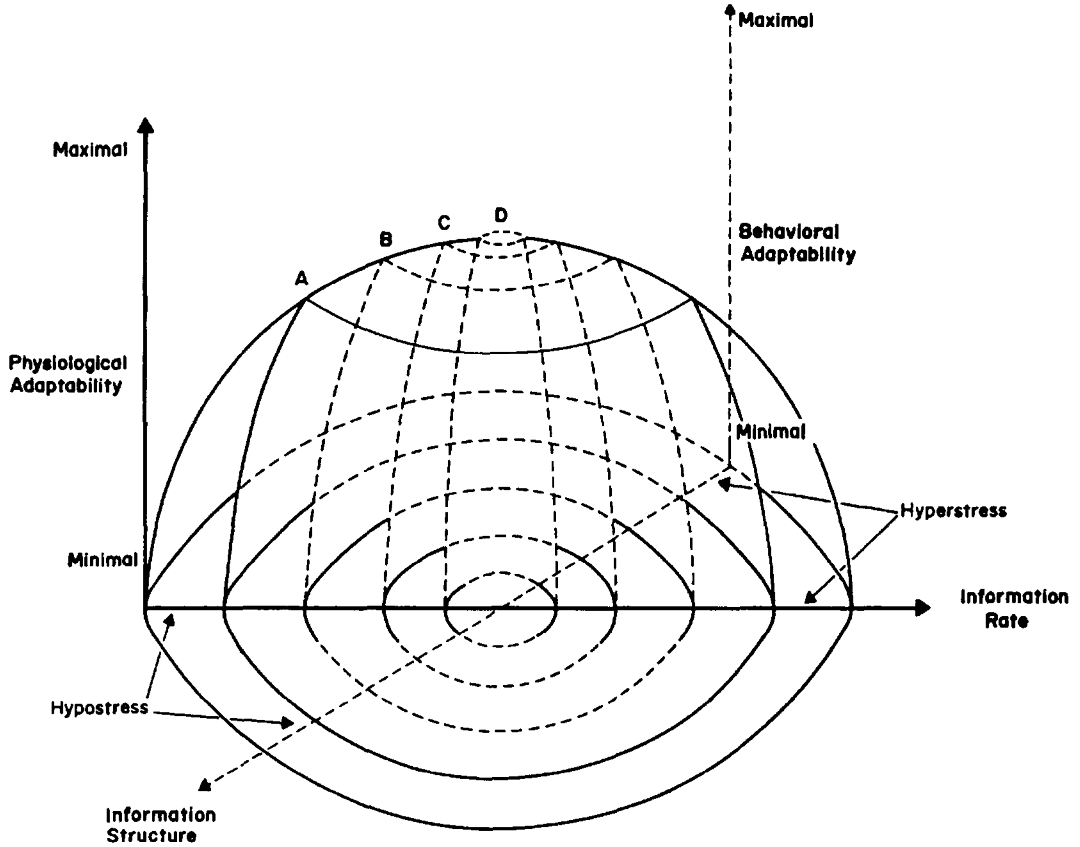

Fig. 1. Physiological and psychological adaptability as functions of hypostress and hyperstress expressed on the dual axes of information rate and information structure. Multiple stresses can be represented as summated scalars plotted as a vector on the two-dimensional base. The necessity for costly adaptive response can be minimized by behavioral strategies which navigate the overall manifold around the perturbations presented by the environment and so avoid stressful conditions. Differing boundaries on the surface of the manifold, A, B, C and D, represent limits to successful capability in some different behavioral categories as indicated in Hancock and Warm (1989).

fatigue. However, it is the case that the cognitive system is, in almost all cases, affected prior to substantial physiological disturbance (Hancock et al., 1990). Consequently, the first thing we see at the onset of behavioral fatigue is adjustments of the cognitive system in seeking varietal relief from the repeated and enforced stimulation. For the long-distance commercial driver, this can mean switching on the radio (augmenting variety through the unused auditory channel), singing to themselves (augmenting output variety through the use of voice) and a variety of other strategies at which the human has proved to be brilliantly facile.

Eventually, it is the input repetition and the enforced residence in those conditions that begins to break down adaptive defenses. Although it can be argued that the commercial driver can take rest breaks, the enforcement comes with the time schedule. Adaptive defenses are remarkable resilient and with context-specific practice (experience in long-distance driving), they can withstand a tremendous degree of punishment. Indeed, we do not see accidents on most long-haul runs. However, it is when a sequence of additional forms of unexpected input demand conspire to push the individual beyond their comfort zone for extended periods that catastrophic performance breakdown occurs. Even then, an accident is not always the result. How many of us have 'woken up' on the freeway only to pull over and give thanks for open-loop control? It is a chain of events, a Markov process, which finally results in the critical mistake. The central questions of remediation is identifying each of the sequential links in that Markov process and fracturing the sequence early in its evolution.

FATIGUE AND WORKLOAD

We propose here that fatigue and workload are reflections of human activity which are of the same universe of discourse. They are each forms of 'energetic' response (Freeman, 1948). Traditionally, fatigue has been associated with extended and repeated operations which imply low and unchanging levels of stimulation. In contrast, much of the research on workload has been directed to an examination of performance under very high levels of demand. Since the vast majority of workload investigations have been undertaken in aviation, this emphasis on high levels of demand is understandable. However, there is now considerable and increasing concern over the workload experienced in 'apparently' low demand situations. The program of work by Warm and his colleagues (see Warm et al., 1996), has demonstrated that extended, obligatory vigilance tasks, which characterize extended driving for example, are not the underloaded situation previously thought. Rather, in accord with the prediction of Hancock and Warm (1989), they are tasks of considerably high workload and the problems associated with fatigue may well be that of continuous high levels of attention demand. Having to pay attention to a display without respite for an extended period, looking for minor variations in an already minimalistic display actually proves very stressful. Warm and his colleagues have also shown that level of workload experienced increases as a function of the length of the work period, consequently, the problems of fatigue are liable to grow non-proportionally with each additional period of work. In sum, we believe that fatigue is directly related to the workload of sustained attention and while differing contributory factors can be involved in generating the fatigue state, the output of such fatigue is impairment akin to the prolonged high workload situation. Given this parallel, we can therefore seek some answers to the problems of driving fatigue in realms which have sought to understand performance in high workload conditions. One of these approaches has been through the use of adaptive systems.

WORKLOAD AND ADAPTIVE SYSTEMS

The aim of adaptive systems is to improve the performance of the entire human–machine system as compared to those static systems which do not adapt. In an ideal situation, an adaptive system creates the optimal environment for the human operator to work in. One of the major aims of adaptive systems is that they prevent the human operator from being overloaded by information and from becoming drowsy due to a lack of information. There has been a lot of work concerning adaptive systems, but there remain questions about their real-world feasibility. We argue here that they are feasible and especially applicable to the fatigued driver problem. This is accomplished by illustrating descriptions of two systems that both are aimed at adapting to the capacities of the driver of a car. One has actually been built and tested, the other is still under development.

It has to be acknowledged that there are inherent problems associated with the development of adaptive systems which adapt dynamically to certain aspects in the environment – as opposed to static 'adaptivity' which indicates a human oriented systems design that does not change over time (see Hancock and Scallen, 1996). The system developer is faced with designing for humans who themselves often respond in unexpected ways because they themselves are adaptive. In terms of control theory, the system developer has to deal with a process of which they can never be sure which parameters can be used to base transmission functions on. In the extreme, this might result in unstable system performance because one process keeps adapting to the other and vice versa. Any designer of adaptive systems, should be aware of this possibility. So, from a systems-engineering perspective, the basic problem of adaptive systems lies in the unpredictable nature of the human component. From a human factors outlook, the problem is to determine how and when the system should adapt to the human operator without counteracting or contradicting the 'normal' human adaptive response. Here, we will deal with system adaptation from the human factors approach: that is, how should a system adapt to the capacities of the human operator?

The notion that human beings adapt to changes in their environment is not new. It is one of the major reasons for human beings having been so 'successful' in their natural environment. In traffic safety, human adaptivity has been advanced as an explanation why certain safety measures are not as successful as they were initially expected to be (see Wilde, 1982; but see also Evans, 1991). This has led some to the idea that humans, either as individuals or as a group, will always adapt perfectly to their environment. Consequently, any change in the environment would be compensated by the human component of the human–machine system. There are indeed indications for complete human adaptation to changes in the environment (e.g. Wilde, 1982) but there are further indications that this principle does not always apply (see, e.g. Evans, 1991). If humans always adapted perfectly there would be no need for augmented technical support in the first place. Humans certainly have many limitations, especially in relation to maladaptive information loads (see Warm et al., 1996).

There is a close relationship between the adaptivity of any system and the workload experienced by the human operator. One of the major reasons to build adaptive systems is to prevent operators from being overloaded by information or, in contrast, to become bored, drowsy and subsequently fatigued. In the area of driving there are efforts to actually build such systems. This implies that the adaptive system needs to, (i) know the state of the driver, and (ii) take over certain tasks which are normally carried out by the driver. Two projects are described in more detail below which aim at developing adaptive systems for driving. The first project is the GIDS project which prevents the driver from becoming overloaded. The second is the SAVE project which aims at reducing accidents due to driver breakdown as caused by, among others, fatigue. In the latter case, the system adapts in that it takes over control if the driver does not perform properly and fails to respond to warnings.

DRIVER OVERLOAD

A recent effort to develop an adaptive in-car system is reported in Michon (1993). This book describes the results of the three-year GIDS (Generic Intelligent Driver Support) project which was funded in part by the European DRIVE program and in which 13 partners from six countries in the European Union participated. Michon gives a detailed account of the behavioral research and technical developments that led to two functional versions of the GIDS system. One of these systems was built into an instrumented vehicle at the TNO Human Factors Research Institute, the other in a fixed base driving simulator at the Traffic Research Centre at the University of Groningen.

The basic notion underlying the GIDS project was that in future, drivers will be confronted with a number of telematics applications in her or his car, some of which aimed at supporting the driver (route guidance, anti-collision, etc.) while others are for communication or entertainment (radio, telephone, fax). These applications can attract the attention of the driver by presenting more or less conspicuous information. One might consider what happens when a driver is negotiating a difficult turn while the route guidance system presents vocal information, the anti-collision system produces a counterforce on the gas pedal, and the phone starts to ring. This might be an infrequent occurrence but that might make the problems even greater when they do occur.

A major aim of the GIDS system is to prevent the driver from becoming overloaded. This is done in three different ways. First, the system is designed such that it makes full use of the human capacity to process information from a variety of sources via different sensory systems. For example, a warning that the vehicle is crossing the line is presented more effectively by means of a slight torque at the wheel as compared to a vocal message. The torque is easier to translate into an appropriate action and interferes less with other auditory messages. This is in the vein of classical human factors issues of distributed attention (e.g. Wickens, 1984) and not directly adaptive in a dynamical way. Second, the GIDS system prevents driver overload by scheduling the information that is presented to the driver. Third, the workload of the driver is examined when the system presents information to the driver. The latter two mechanisms are truly adaptive and are described in more detail below.

SCHEDULING IN GIDS

Scheduling is the process which ensures that the driver is not overloaded by various in-car messages that are presented in close temporal proximity. The GIDS architecture is designed such that no application can present its information directly to the driver. Instead, it transfers the information to a central core system, the Scheduler. Along with a symbolic description of the message content, each application indicates to the Scheduler the output device (display, speaker, gas pedal, or steering wheel), the priority of the message (in terms of traffic safety) and the time required for the driver to process the message. It is the Scheduler's task to prioritize these messages. Given that the priority of a message changes over time, such as a route guidance message which might get more important as the vehicle approaches the intersection, the priority within the Scheduler changes continuously. After presentation of each message, the Scheduler waits to allow the driver to process the message and take the necessary actions.

The Scheduler mechanism takes care that two messages are never presented at the same time. In other words, it considers the human as a single channel processor. As humans are clearly able to do more than one thing at a time, future versions of GIDS-like systems should include mechanisms to allow certain messages to be displayed concurrently. One possible solution would be to have each potential message accompanied by an indication of how great the mental workload (recognition and decision) of the message imposes and the extent it can be generated at the same time as messages in other sensory modalities. In this way the system would adapt to the human ability to process information concurrently when different sensory modalities are being used and the information is highly familiar (e.g. Wickens, 1984).

In this latter, more sophisticated, form the Scheduler represents a full example of both static and dynamic adaptation. Static in that it recognizes the human problem to process unfamiliar information in the same sensory modality at the same time, dynamic in that it is also able to acknowledge the human capacity to process familiar sorts of information in different sensory channels concurrently.

WORKLOAD ESTIMATION IN GIDS

Scheduling might prevent the driver from being overloaded by simultaneous presentation of two or more messages. It does not take the workload into account caused by the driving task itself. Obviously, this is an important issue since workload can vary according to driving context. The basic problem here is how a system can know what the driver's current level of workload is. On-line measurement is hard in a normal car (even though steering behavior might give a clue – Verwey and Veltman, in press). The alternative is to estimate driver workload on the basis of a simple model. In a series of experiments, some major determinants of driver workload were determined (Verwey, 1993). Furthermore, a distinction was made between mental and visual workload, as these are the types of workload that are most likely to become overloaded. These studies showed that the major determinant of driver workload is the driving situation. Furthermore, mental workload was relatively high for inexperienced and elderly drivers in more complex situations such as turns and roundabouts. Time of day and familiarity with the environment (when guided by a proper route guidance system) contributed little to the variations in driver workload.

On basis of these results the Workload Estimator module was built which provides estimates of driver workload by way of a look-up table. With the estimates of the Workload Estimator, the Scheduler can decide whether to present complex messages of a lower priority immediately, or postpone them until the Workload Estimator indicates lower workload levels again. Preliminary results with experienced, younger and middle-aged drivers show that the Workload Estimator did a good job with respect to visual workload but a much poorer estimate with respect to mental workload (Verwey, in prep.). Given that variations in mental workload are limited for these drivers and not for inexperienced drivers, mental workload estimation might still do a good job for the latter novice group of drivers.

DRIVER UNDERLOAD

As discussed at length at many places in the literature, attentional failures are a common cause of accidents. Such failures may be caused by momentary, transient distractions and periods of overload. In addition, they also occur due to driver fatigue (e.g. Knipling and Wang, 1995). With the development of systems such as intelligent cruise controls, the chance of driver drowsiness may even increase (Verwey et al., 1996; Hogema et al., 1996).

Whereas the detection of driver overload is not easy, reliable detection of reduced attentional states poses even greater problems. Over ten years ago, Zaidel (1985) mentioned the three basic problems with respect to driver monitoring: (a) what should be measured? (b) what are the criteria for activating a warning? and (c) what is the appropriate form of warning? We believe that new technologies allow us to take another step forward in the development of a system capable of detecting driver breakdown and taking action when necessary. This section describes a second three year research project (SAVE) that is carried out within the framework of the European Union's Transport Telematics Program. The consortium that carries out the project includes 10 partners from seven countries within the European Union. As the project started approximately six months ago, results of behavioral studies are not presented here. However, we do provide an overview of the philosophy behind the project and the resulting system architecture.

SAVE SYSTEM ARCHITECTURE

The SAVE application consists of three components or sub-applications. Their mutual relationship is depicted in Fig. 2. The Integrated Monitoring Unit (IMU), the Automatic Control Device (ACD) and the Save Warning System (SWS). Each is discussed briefly below.

The Integrated Monitoring Unit (IMU) is directed to detecting that the driver is unable to drive safely. Given the different reasons for unsafe driving, associated with different behavioral patterns, the IMU consists of three further subsystems. The Driver Drowsiness Detector (DDD) determines whether the driver is about to fall asleep, the Alcohol Detector (ALD) recognizes that the driver is intoxicated and the Brake Down Detector (BDD) identifies that the driver is not undertaking actions any longer due to circumstances like health problems, vasovagal collapse, epileptic attacks, hyperventilation, fear attacks, etc.

If the diagnosis is made that driving safety is jeopardized due to the driver's present state, the SAVE application undertakes action. Two sub-applications are responsible for these actions. The Automatic Control Device (ACD) will stop the car safely alongside the road when the driver is not

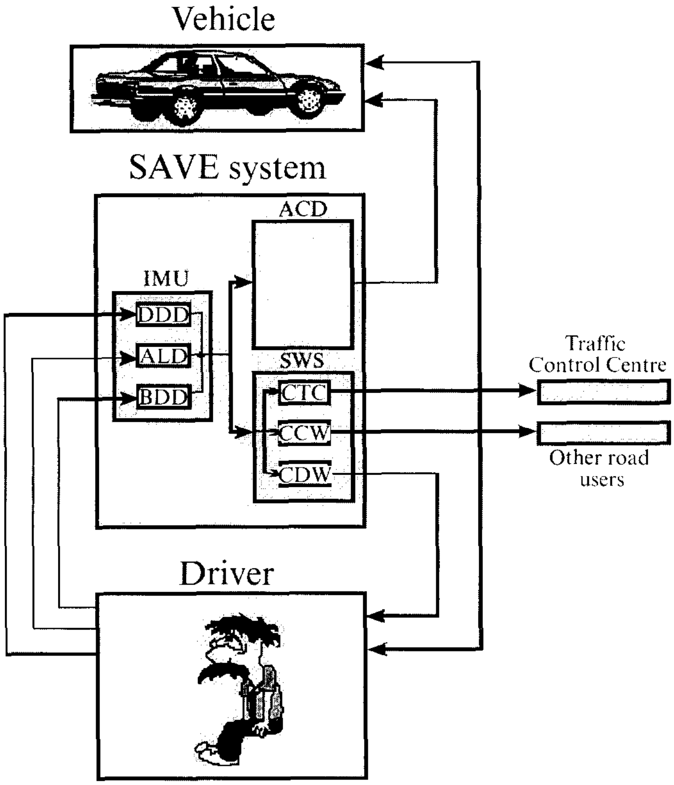

Fig. 2. Relationship amongst the various components, the driver, the vehicle and the outside world within the SAVE application.

responding at all. This requires no human involvement and is carried out on the basis of the information provided by sensors indicating the presence of other vehicles and relevant objects and the car's current lane position.

If the IMU detects that the driver is not acting safely but may still be able to stop the vehicle, the Save Warning System (SWS) will first warn the driver in an appropriate way via the Car to Driver Warning (CDW) system. At the same time, or after the car has been stopped, the Car to Car Warning (CCW) system informs other drivers about the fact that the car has been stopped due to possible health problems of its driver. Finally, the traffic control center is informed through the Car to Traffic Control (CTC) system.

So, the SAVE application involves three major components. (1) The IMU is responsible for detecting that the driver is unable to drive safely. This might be caused by drowsiness and fatigue (DDD), alcohol (ALD) and total break down (BDD). (2) The SWS will, depending on the situation at hand, warn the driver (CDW), the immediate environment of the vehicle (CCW) and/or the traffic control center (CTC). (3) The ACD will take care that, in case the driver does not respond properly, the vehicle stops safely.

SAVE RESEARCH OBJECTIVES

The goal of SAVE is to develop an operational system, capable of detecting driver breakdown due to fatigue, alcohol or medical problems and take appropriate action. It is clear that many problems are to be solved. These relate to driver break down detection but also to the actions to be taken by the system.

With respect to the detection of driver breakdown, a multi-sensor approach has been chosen. Earlier research has shown that various indicators can be used for determining driver break down (e.g. Mackie and Wylie, 1991). However, there is no single indicator that can do the job reliably. Furthermore, several of the more reliable indicators are physiological and can not be used other than as a reference for drowsiness in experimental conditions. Basically, the approach taken defines a core set of behavioral parameters which are to be measured in a large number of studies in which drivers will drive until they fall asleep or when they are alcohol intoxicated. This core set will be 'fed' to a neural net along with a signal indicating when driving safety is no longer acceptable. This core set will consist of speed variations, headway, steering wheel movements and steering wheel grip force, lateral position of the vehicle, eye blink rate and eye closure time, and head position.

Earlier work has identified absolute criteria for unsafe driving behavior (Verwey et al., 1996). Besides, an attempt will be made to decide upon safety by relating behavior under illegal levels of blood alcohol with driving behavior in drowsy conditions. Apart from this driver monitoring strategy, research will also investigate the role of certain static variables such as time of the day, time of driving, sleep deprivation, driving environment, climate and individual characteristics (including health status and personality-related factors). For example, driving in a familiar environment seems to make drivers more drowsy than an unfamiliar environment (Wertheim, 1991) and extroverted and old and young people seem more vulnerable to fatigue than introverted and middle-aged drivers (Berch and Kanter, 1984; Horne and Reyner, 1995). These types of data will also be input to the neural net. Depending on the data used to train the neural net, the system may learn to recognize breakdown of a particular individual or of an average driver. Preferably, it will initially be trained for the average driver and, next, it may learn the peculiarities of 'its' own driver and how that driver performs in different conditions. In other words, the system will adapt to the regular driver of the car. If several people drive the car this should obviously be indicated before individual departure.

Another issue concern the actions to be taken given that the driver breaks down. First of all, the SAVE system will not only detect that the driver breaks down, but also whether they are still conscious and able to stop. In that case, a warning will be issued to stop as soon as possible. If the driver is not capable of undertaking any action, the ACD will warn people in the nearest Traffic Control Center and in the immediate traffic environment and automatically stop the vehicle at the shoulder. Apart from total unconsciousness, the general philosophy is that the driver remains responsible and the system will not take over. Special consideration will be given to the way in which the driver is being warned. It has been shown various times that when a drowsy driver is being warned because the vehicle is leaving its lane, the driver startles and tends to oversteer. This indicates that the system should come up with a warning before the vehicle is leaving its lane. In addition, however, research will investigate the possibility to warn the driver by temporarily taking over the steering task until the vehicle has either stopped automatically, or it is back in its lane again.

On basis of the results obtained in the first two years, parameters of the SAVE system will be determined, components will be built and tested in isolation. In the third year, the components will be assembled and the integrated SAVE system will be put to test under simulated and real road driving conditions. Even in relatively complex driving environments and diverse driving populations, adaptive systems are feasible. The major goal of these systems is to protect the driver from extreme levels of workload. Situations of information overload can be prevented by scheduling the various messages from different in-car telematics applications and by postponing or cancellation of low priority messages in complex driving situations. With respect to driving on long and boring stretches of road, an adaptive system should continuously monitor the driver and determine her or his current state, warn if possible, and take over in case the driver is no longer able to control the car properly. In the latter case, the adaptive property of the system lies primarily in the possibility to take over.

In contrast to adaptive systems in professional environments, such as in airplane cockpits and operator rooms, adaptive systems in the in-car environment have to deal with a large variability of human capacities. In fact, this implies that the system should adapt not only to the environmental conditions but, more than in professional environments, it should adapt to the idiosyncrasies of the human operator. This could be done by making use of the learning capabilities of modern system components (e.g. neural nets) but presetting the system to the capacities of the current driver seems more feasible on the shorter run (Verwey et al., 1993). Given the wide variety of human capacities, a distinction should be made on the basis of group distinctions such as age and experience. Finally, the system might have the human operator execute a performance test in order to classify the human operator.

HUMAN-CENTERED ADVANCED DRIVER SYSTEMS

The previous examples are illustrations of what we have termed human-centered transportation systems. The systems under development in the GIDS and SAVE projects have one goal: improving the quality of the behavior of the driver-car system by designing a human-centered system. Current and future technology will allow the systems designer to develop increasingly intelligent systems. However, this certainly does not guarantee that the systems are designed intelligently. Given the increasing complexity of modern systems the need for human-centered system design is increasing radically (see Hancock and Scallen, 1996). The typical engineering approach to systems design reasons according to the adage 'if I can control it, everybody can' but will result, now more than ever, in inadequate system design from a human factors perspective. Humans are adaptive themselves, and are able to deal with all kinds of poorly designed systems, but this does not mean that they will perform well under all circumstances. The two earlier examples show precisely when problems are expected with systems that have not been developed from a human-centered design perspective, namely in situations of underload and of overload.

The ubiquitous phenomenon of stimulus-response compatibility may serve as an example: Everybody can learn to push a button at the left in response to a light being switched on at the right. And from an engineering point of view there may be no reason to worry about the slight increase in reaction time. Yet, it has been demonstrated many times that it is not so much the increase in reaction time but the increase in error rate under situations of over- and underload that is worrying. Operators who are either extremely busy or highly fatigued because they have been on duty for a long time, simply tend to fall back on the innate compatibility principle of responding in the direction of the stimulus.

It is important to realize that the need for human centered systems design is perhaps even more urgent for the driving environment (the car) than for systems in aircraft and operating rooms. While training of highly skilled professionals may compensate for poor systems design, this will certainly not be the case with in-car systems which will not only be used by the highly educated, who know about technology, but also by people who do not know, and often don't want to know about technology, and people with limited (mental and physical) abilities. Even without advanced in-car systems, much effort has been directed to understanding why accident rates of elderly and inexperienced drivers are so high. It is highly likely that poor design of future in-car systems will only worsen the situation.

Given the increasing degrees of freedom for system designers and the increasing demand for human flexibility to control complex systems, the need for human centered systems design increases too. Traditional knobs/dials and display-color ergonomics represent contemporary approaches to adapting systems to humans. Future human-centered systems design will involve various types of adaptation. First of all, human centered systems design will have to rely on static adaptation. Static adaptation indicates that systems design follows the guidelines provided by, what could be called, traditional ergonomics. This implies that the system is based on what we know about human capabilities and capacities in general but once incorporated in the design, the system's behavior will not change any more. An illustration in car control design is the design of most modern dashboard panels and the conformity of typical car control locations and operational characteristics of accelerator and brake pedals amongst different types of cars. One could say that cars are designed according to these principles because people are found to be able to handle them well. However, various deviations from these ergonomic principles can still be found, such as with digital speedometers in some cars.

A more complex type of human centered systems design, is dynamic adaptation. Dynamic adaptation can be considered the type of adaptation which, together, takes human capacities into account by changing aspects of the system's behavior while it is in use. Two forms of dynamic adaptation can be distinguished, generalized and idiosyncratic adaptation. Generalized adaptation indicates dynamic adaptable systems which take into account the effects of external (input) variables for an average operator or driver. The development of the GIDS system was largely based on the notion of generalized dynamic adaptation: humans in general are not able to process two visual messages at the same time and they are not able to watch a map display while negotiating a roundabout. Therefore, the system should prevent such messages from being presented in those situations. On the other end of the workload continuum, the SAVE system monitors the driver and will either warn the driver or take over car control and stop the car if the driver is no longer able to do so.

Idiosyncratic dynamic adaptation goes one step further in that the system takes the capacities of the current operator, or the current driver into account. This requires the system to be adapted to the specific user. Current technology allows two ways to achieve this adaptation. Either the system requires the user (or someone else) to define the required behavior in various situations. This could be termed human-initiated calibration (Verwey et al., 1993). Alternatively, the system is able to learn about the idiosyncrasies of the user. The latter is possible only when the system is able to interpret the user's behavior and know when poor behavior occurs. This can be called system-initiated calibration.

Consequently, human-centered systems design involves (a) static adaptation, (b) generalized dynamic adaptation and (c) idiosyncratic dynamic adaptation. The latter type of adaptation can be accomplished either by the human (the user or someone else) or by the system. At this moment, car design is largely based on static adaptation. Yet, efforts are currently underway to develop dynamic adaptation in the car environment. Up till now, these efforts are largely aimed at generalized dynamic adaptation: the in-car system adapts to the capacities of drivers in general. However, in some projects (such as GIDS and SAVE), attempts have already been made to develop idiosyncratic adaptation in terms of adapting the system's behavior to the type of driver. This is achieved by informing the system about the type of driver, a form of human-initiated calibration. As our knowledge about measuring and interpreting driver behavior progresses along with systems' computational capabilities, system-initiated calibration will also become a reality.

SUMMARY AND CONCLUSIONS

As a multi-dimensional affective state, fatigue still remains a difficult attribute to define and measure unequivocally. This should not be terminally discouraging, since many comparable affective states can and inevitable do suffer from the same drawback[4]. However, the necessity to understand fatigue is driven not merely by intrinsic scientific curiosity but the very real need of the operational community to mitigate fatigue effects in their operations, where its effects can be fatal. We have suggested here that fatigue can be directly related to stress and great advantage accrues from this cross-fertilization. By extension, we can link workload to stress and thus provide quantitative inputs to systems purposely-designed to promote safe driving. The adaptive systems we have cited do not reduce or eliminate fatigue per se. Rather, they seek to defend the driver from the adverse consequences that fatigue can induce.

The answer to fatigue is actually very simple – provide sufficient rest. It is an answer that society and individuals within society often do not want to hear. It comes down ultimately to what we believe the purposes of technology are, and the price we are willing to pay to achieve those purposes. At present, in respect of fatigue, the price is too high.

REFERENCES

Berch, D.B. and Kanter, D.R. (1984) Individual differences. In: Warm, J.S., ed. *Sustained Attention in Human Performance*. pp. 143–177. John Wiley, Chichester.

Broadbent, D.E. (1979) Is a fatigue test now possible? Ergonomics **22**, 1227–1290.

Evans, L. (1991) *Traffic Safety and the Driver*. Van Nostrand Reinhold, New York.

Freeman, G.L. (1948) *The Energetics of Human Behavior*. Ithaca Press, Cornell University Press.

Hancock, P.A. (1980) Sensory deprivation in long distance swimming: A psychological stress factor?. *Swimming Technique* **16**, 122–123.

Hancock, P.A., Chignell, M.H. and Vercruyssen, M. (1990) Predicting the effects of stress on performance. *Proceedings of the Human Factors Society* **34**, 1081–1085.

Hancock, P.A. and Scallen, S.F. (1996) The future of function allocation. *Ergonomics in Design*. 4(4), 24–29.

Hancock, P.A. and Warm, J.S. (1989) A dynamic model of stress and sustained attention. *Human Factors* **31**, 519–537.

Hogema, J.H., Janssen, W.H., Coemet, M. and Soeteman, H.J. (1996) Effects of intelligent cruise control on driving behaviour: A simulator study. Paper No. 2081. *Proceedings of the Third Annual Congress on Intelligent Transport Systems*.

Horne, J.A. and Reyner, L.A. (1995) Sleep related vehicle accidents. *British Medical Journal* **310**, 565–567.

Kant, I. (1787) *The Critique of Pure Reason*. Hartknocht, Riga.

Knipling, R.R. and Wang, J.S. (1995) Revised estimates of the U.S. drowsy driver crash problem size based on general estimates systems case reviews. *Proceedings of the 39th Annual Association for the Advancement of Automotive Medicine*.

Mackie, R. and Wylie, C.D. (1991) Countermeasures to loss of alertness in motor vehicle drivers: A taxonomy and evaluation. *Proceedings of the 35th Annual meeting of the Human Factors Society*, pp. 1149–1153, Human Factors Society, Santa Monica, CA.

Michon, J.A., ed. (1993) *Generic Intelligent Driver Support, A Comprehensive Report on GIDS*. Taylor and Francis, London.

Monk, T. and Folkard, S. (1983) Circadian rhythms and shiftwork. In: *Stress and Fatigue in Human Performance*, ed. G.R.J. Hockey, pp. 97–121. Wiley, New York.

Muscio, B. (1921) Is a fatigue test possible?. *British Journal of Psychology* **12**, 31–46.

Verwey, W.B. (1993) How can we prevent overload of the driver? In: *Driving Future Vehicles*, eds. A.M. Parkes, S. Franzen, pp. 235–244. Taylor and Francis, London.

Verwey, W.B. (1997) On-line driver workload estimation: Towards an adaptive human machine interface in the car of tomorrow. in revision, (in preparation).

Verwey, W.B., Alm, H., Groeger, J., Janssen, W.H., Kuiken, M.J., Schraagen, J.M., Schumann, J., van Winsum, W. and Wontorra, H. (1993) GIDS functions. In: *Generic Intelligent Driver Support. A Comprehensive Report on GIDS*, ed. J.A. Michon, pp. 113–146. Taylor and Francis, London.

Verwey, W.B., Brookhuis, K.A. and Janssen, W.H. (1996) Safety effects of in-vehicle information systems. (Report TM-96-C002). Soesterberg, The Netherlands: TNO Human Factors Research Institute.

Verwey, W.B. and Veltman, J.A. (1997) Measuring workload peaks while driving. A comparison of nine common workload assessment techniques. *Journal of Experiment Psychology: Applied* (in press).

Warm, J.S., Dember, W.N. and Hancock, P.A. (1996) Vigilance and workload in automated systems. In: *Automation and Human Performance: Theory and Applica-*

[4]It can well be argued that no semantic or mathematical description of an affective state or emotion can ever capture the essence of what is trying to be defined. Quite simply, words and numbers are representative of a symbolic description of a state of an entity which is felt rather than described. In this respect, fatigue falls alongside all such constructs as affection, anger, jealousy and love, which poets have been trying to capture for millenia. They have yet to succeed although some have gotten closer than others.

tions, eds. R. Parasuraman, M. Mouloua. Erlbaum, Hillsdale, N.J.

Wertheim, A.H. (1991) Highway hypnosis: A theoretical analysis. In: *Vision in Vehicles III*, eds. A.G. Gale, I.D. Brown, C.M. Haslegrave, I. Moorhead, S. Taylor, pp. 467–472. North-Holland, Amsterdam.

Wickens, C.D. (1984) Processing resources in attention. In: *Varieties of Attention*, eds. R. Parasuraman, D.R. Davies, pp. 63–102. Academic Press, London.

Wilde, G.J.S. (1982) The theory of risk-homeostasis: implications for safety and health. *Risk Analysis* **2**, 209–255.

Zaidel, D. (1985) Driver alerting devices – new technology, old problems. In: *Ergonomics International 85, Proceedings of the Ninth Congress of the International Ergonomics Association*, eds. I.D. Brown, R. Goldsmith, K. Coombes, M.A. Sinclair, pp. 436–438. Taylor and Francis, London.